Data Transport

(see Appendix 2 for a more detailed discussion)

The concept of a "web service" is fundamental to the ability of DMAC to connect quasi-independent systems. The term "web service" is used in many contexts today; in the DMAC Plan we intend the term to mean reusable software components that provide a standardized means for computer systems to request data and data processing from one another, typically using messages expressed in the eXtensible Markup Language (XML) and conveyed using the ubiquitous communications protocol of the World Wide Web, HTTP (the Hyper Text Transfer Protocol). Web services make data and software capabilities available on one computer, accessible to other computers via the Internet through the familiar Universal Resource Identifiers (URIs) that begin with "http://".

DMAC will endorse a suite of web services to serve as a shared communications toolbox connecting systems that are operated by regional, state, and federal agencies; academic projects; international partners; and others. Data suppliers (including PDA&QC sites) will be responsible for making data accessible through DMAC web services tools and standards. Data users will find that in many cases the software applications upon which they depend for product generation and scientific analysis will be "DMAC ready" (possibly with some adaptation required), having been adapted to work directly with the DMAC web services. In this case, the applications will perform much as if the data existed on users' local hard drives. To provide a bridge from current practices, compatibility between DMAC and user applications may be achieved using formatted files that are made readily available as products through DMAC web services.

The DMAC data transport framework will designate a suite of freely available software components adequate to meet typical needs. The goal in doing so is to minimize the barriers to participation in the DMAC. The uniformity provided by the DMAC web services standards will permit all the components related to data transport to be interoperable at the machine level (i.e., data can be moved from one component of the system to another, retaining complete syntactic[9] and semantic[10] meaning without human interaction).

[9]Syntactic Meaning: refers to the syntax of a data set—the atomic data types in the data set (e.g., binary, ASCII, real), the dimensionality of data arrays (P is a 90 by 180 by 25 by 12 element array), the relationship between variables in the data set (lat is a map vector for the first dimension of P), etc.

[10]Semantic Meaning: refers to the semantics of the data contained in the data set—the meaning of variables (P represents phytoplankton abundance), the units used to express variables (multiply P by 8 to obtain number of specimens per cubic meter), special value flags (a value of –1 means missing data, 0 land,...), descriptions of the processing or instrumentation used to obtain the data values, etc.

Web services exist in the context of the web. Data transport on the web involves protocols at multiple levels. The foundation of transport on the Internet is TCP/IP, which handles the routing of "packets" of information between source and destination hosts. Layered upon TCP/IP are a variety of protocols, for example, FTP, HTTP[11], and SMTP. These protocols are supported on a very wide range of computers and operating systems, and all of them will be used to move various types of data over the network as part of IOOS. There is, however, no uniform syntactic and semantic meaning that is guaranteed for data communicated via these transfers, and therefore no guarantee of immediate interoperability among computer applications. This function is the role of the DMAC web services standards.

Several solutions currently exist for the syntactic description and transport of binary data, however none is universally accepted. The most broadly tested and accepted of these solutions within oceanography are the OPeNDAP[12] data access protocol and the Open Geospatial Consortium, Inc. (OGC) (http://www.opengeospatial.org/) data access protocols.

OPeNDAP

OPeNDAP underlies the National Virtual Ocean Data System (NVODS[13]). OPeNDAP has been serving the marine community since 1995. OPeNDAP provides the very general approach to data management that is needed in support of research and modeling. OPeNDAP also supports server-side subsetting of data, which greatly reduces the volumes of data that need be transferred across the Internet in many cases. This capability is vital when considering the large volumes of data that will be produced in the near future by observing platforms and modeling activities. Tables 2 and 3 provide estimates of near-term data flow for the U.S. IOOS using selected data streams as examples; the lists are not intended to include all observing system data types. The DMAC Steering Committee recommends the designation of OPeNDAP protocol as an initial "operational"[14] component for Data Transport of gridded data, and a "pilot"[15] component for the delivery of non-gridded data.

[11]When using a web browser, most images and text are delivered via http.

[12]The Open Source Project for a Network Data Access Protocol (OPeNDAP) is a non-profit corporation formed to develop and maintain the middleware formerly known as the Distributed Ocean Data System (DODS).

[13]NVODS was created in response to a Broad Agency Announcement (BAA) issued by the National Oceanographic Partnership Program in 2000.

[14]"Operational" is stage four of a four-level classification scheme for the maturity of system components within IOOS: R&D, pilot, pre-operational, operational. See IOOS Development Plan (www.ocean.us).

[15]"Pilot" is stage two of a four-level classification scheme for the maturity of system components within IOOS: R&D, pilot, pre-operational, operational. See IOOS Development Plan (www.ocean.us).